Matrix Approach to Linear Regression

Frank Wood

March 2, 2010
Random Vectors and Matrices

Let’s say we have a vector consisting of three random variables

\[ \mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \]

The expectation of a random vector is defined as

\[ \mathbb{E}(\mathbf{y}) = \begin{pmatrix} \mathbb{E}(Y_1) \\ \mathbb{E}(Y_2) \\ \mathbb{E}(Y_3) \end{pmatrix} \]
Expectation of a Random Matrix

The expectation of a random matrix is defined similarly

\[ \mathbb{E}(y) = [\mathbb{E}(Y_{ij})] \quad i = 1, \ldots, n; j = 1, \ldots, p \]
Covariance Matrix of a Random Vector

The correlation of variances and covariances of and between the elements of a random vector can be collection into a matrix called the covariance matrix

$$\text{cov}(\mathbf{y}) = \sigma^2\{\mathbf{y}\} = \begin{pmatrix}
\sigma^2(Y_1) & \sigma(Y_1, Y_2) & \sigma(Y_1, Y_3) \\
\sigma(Y_2, Y_1) & \sigma^2(Y_2) & \sigma(Y_2, Y_3) \\
\sigma(Y_3, Y_1) & \sigma(Y_3, Y_2) & \sigma^2(Y_3)
\end{pmatrix}$$

remember $\sigma(Y_2, Y_1) = \sigma(Y_1, Y_2)$ so the covariance matrix is symmetric
Derivation of Covariance Matrix

In vector terms the covariance matrix is defined by

$$\sigma^2\{y\} = \mathbb{E}(y - \mathbb{E}(y))(y - \mathbb{E}(y))^\prime$$

because

$$\sigma^2\{y\} = \mathbb{E}\left( \begin{pmatrix} Y_1 - \mathbb{E}(Y_1) \\ Y_2 - \mathbb{E}(Y_2) \\ Y_3 - \mathbb{E}(Y_3) \end{pmatrix} \right) \begin{pmatrix} Y_1 - \mathbb{E}(Y_1) & Y_2 - \mathbb{E}(Y_2) & Y_3 - \mathbb{E}(Y_3) \end{pmatrix}$$
Regression Example

- Take a regression example with $n = 3$ with constant error terms $\sigma^2(\epsilon_i)$ and are uncorrelated so that $\sigma^2(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$
- The covariance matrix for the random vector $\epsilon$ is

$$
\sigma^2(\epsilon) = \begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & 0 \\
0 & 0 & \sigma^2
\end{pmatrix}
$$

which can be written as $\sigma^2\{\epsilon\} = \sigma^2 I$
Basic Results

If \( A \) is a constant matrix and \( Y \) is a random matrix then \( W = AY \) is a random matrix

\[
\mathbb{E}(A) = A \\
\mathbb{E}(W) = \mathbb{E}(Ay) = A \mathbb{E}(y) \\
\sigma^2\{W\} = \sigma^2\{Ay\} = A \sigma^2\{y\} A'
\]
Multivariate Normal Density

- Let $Y$ be a vector of $p$ observations

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix}$$

- Let $\mu$ be a vector of $p$ means of each of the $p$ observations

$$Y = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$
Multivariate Normal Density

let $\Sigma$ be the covariance matrix of $Y$

$$
\Sigma = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\
\sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2
\end{pmatrix}
$$

Then the multivariate normal density is given by

$$
P(Y|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(Y - \mu)'\Sigma^{-1}(Y - \mu)\right)
$$
Example 2d Multivariate Normal Distribution
Matrix Simple Linear Regression

- Nothing new-only matrix formalism for previous results
- Remember the normal error regression model
  \[ Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n \]
- Expanded out this looks like
  \[
  Y_1 = \beta_0 + \beta_1 X_1 + \epsilon_1 \\
  Y_2 = \beta_0 + \beta_1 X_2 + \epsilon_2 \\
  \vdots \\
  Y_n = \beta_0 + \beta_1 X_n + \epsilon_n
  \]
- which points towards an obvious matrix formulation.
Regression Matrices

If we identify the following matrices

\[
Y = \begin{pmatrix}
Y_1 \\
Y_2 \\
. \\
. \\
. \\
Y_n
\end{pmatrix}, \quad
X = \begin{pmatrix}
1 & X_1 \\
1 & X_2 \\
. \\
. \\
. \\
1 & X_n
\end{pmatrix}, \quad
\beta = \begin{pmatrix}
\beta_0 \\
\beta_1
\end{pmatrix}, \quad
\epsilon = \begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
. \\
. \\
. \\
\epsilon_n
\end{pmatrix}
\]

We can write the linear regression equations in a compact form

\[y = X\beta + \epsilon\]
Of course, in the normal regression model the expected value of each of the $\epsilon$’s is zero, we can write $\mathbb{E}(y) = X\beta$

This is because

$$\mathbb{E}(\epsilon) = 0$$

$$\begin{pmatrix}
\mathbb{E}(\epsilon_1) \\
\mathbb{E}(\epsilon_2) \\
\vdots \\
\mathbb{E}(\epsilon_n)
\end{pmatrix} =
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}$$
Error Covariance

Because the error terms are independent and have constant variance $\sigma^2$

$$\sigma^2\{\epsilon\} = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix}$$

$$\sigma^2\{\epsilon\} = \sigma^2 I$$
Matrix Normal Regression Model

In matrix terms the normal regression model can be written as

\[ y = X\beta + \epsilon \]

where \( \mathbb{E}(\epsilon) = 0 \) and \( \sigma^2\{\epsilon\} = \sigma^2 I \), i.e. \( \epsilon \sim N(0, \sigma^2 I) \)
Least Square Estimation

If we remember both the starting normal equations that we derived

\[ nb_0 + b_1 \sum X_i = \sum Y_i \]
\[ b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i \]

and the fact that

\[
X'X = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
X_1 & X_1 & \cdots & X_n \\
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1
\end{bmatrix}
= \begin{bmatrix}
1 \\
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix}
= \begin{bmatrix}
n \\
\sum X_i \\
\sum X_i^2
\end{bmatrix}
\]

\[
X'y = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
X_1 & X_1 & \cdots & X_n \\
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1
\end{bmatrix}
= \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{bmatrix}
= \begin{bmatrix}
\sum Y_i \\
\sum X_i Y_i
\end{bmatrix}
\]
Least Square Estimation

Then we can see that these equations are equivalent to the following matrix operations

\[ X'X \ b = X'y \]

with

\[ b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \]

with the solution to this equation given by

\[ b = (X'X)^{-1}X'y \]

when \((X'X)^{-1}\) exists.
When does $(X'X)^{-1}$ exist?

$X$ is an $n \times p$ (or $p + 1$ depending on how you define $p$) design matrix.

$X$ must have full column rank in order for the inverse to exist, i.e.

$\text{rank}(X) = p \iff (X'X)^{-1}$ exists.